Composability in a certain family of entropies

Henrik Densing Petersen*
Department of Mathematical Sciences
University of Copenhagen
m03hdp@math.ku.dk

February 1, 2008

Abstract

It is shown that the Tsallis entropies are the only entropies of the form $H(P) = -\sum_i f(p_i)$, with suitable assumptions on f, satisfying the condition of composability.

Keywords. Composability, Tsallis entropy, f-entropy.

1 Introduction

We define the f-entropy as in [1] (see also [2] for some basic properties), namely, we demand that f is a real-valued analytic and strictly convex function on $[0, \infty)$, that it satisfies $f(0) = f(1) = 0$, and lastly, that it is normalized so that $f'(1) = 1$. The f-entropy is then defined as

$$H_f(P) = -\sum_{i \in A} f(p_i), \tag{1}$$

where $P = (p_1, p_2, \dots)$ is a probability distribution on A. Unless explicitly stated otherwise we shall assume that $A = \mathbb{N}$.

Simple considerations show that for any probability distribution P, $0 \leq H_f(P) < -f'(0)$ and that $\sup\{H_f(P)|P\} = -f'(0)$. Thus g in (2) need only be defined on $[0, -f'(0)[\times [0, -f'(0)[$.

The purpose of this paper is to show that the Tsallis entropies are the only f-entropies that are composable, i.e. satisfy the equation

$$H_f(P \otimes Q) = g(H_f(P), H_f(Q)) \tag{2}$$

for some function g and all probability distributions P, Q. The possible forms of the function g for general measures of entropy is considered in the articles

---

*Personal contact information: Henrik Densing Petersen, Melissehaven 23, 2.th., 2730 Herlev, Denmark. Phone: +45 61717642. Fax (at the institute): +45 35320704

[8, 11]. The relation between non-additive and additive entropies is explored in [10].

The Tsallis entropies were first considered in the mathematical literature in [5] and then later (and independently) in statistical physics by C. Tsallis in [3]. They are defined for any $q > 0$ as the f-entropy associated with the function

$$f(x) = \frac{x - x^q}{1 - q}.$$

It is not hard to see that the Tsallis entropies in fact satisfy the equation

$$H_q(P \otimes Q) = H_q(P) + H_q(Q) + (1 - q)H_q(P)H_q(Q). \tag{3}$$

Our first theorem is the converse: Any f-entropy satisfying an equation of this type must be a Tsallis entropy. Having proved that, we can extend the result to cover f-entropies satisfying the more general equation (2).

The paper [6] addresses the same issue as here. Let us briefly describe how the present work differs in outlook. In [6], the "non-obvious" assumptions needed are the existence of $f'(0)$ as well as either that of f being a sum $f(x) = \sum_{\text{some } \alpha's \geq 1} x^\alpha \phi_\alpha(x)$ or an assumption of analyticity on g.

Assuming the existence of $f'(0)$ seems rather limiting, recalling that for the shannon entropy, $(x \log x)'$ indeed does not exist for $x = 0$. As for the other two assumptions, it is not at all a priori clear how restrictive they are in physical terms.

It seems much more natural to take an important physical property of the entropy, in our case concavity (which is equivalent to the convexity of f), and starting from there. This gives a precise result, which seems to have exactly the right relevance for physical applications.

2   Composability of f-entropy

Theorem 1. *If the f-entropy satisfies*

$$H_f(P \otimes Q) = \alpha H_f(P) + \beta H_f(Q) - \gamma H_f(P)H_f(Q), \tag{4}$$

for constants α, β, γ and all probability distributions P, Q, then $\alpha = \beta = 1$, and

$$f(xy) = xf(y) + yf(x) + \gamma f(x)f(y) \tag{5}$$

for all $x, y \geq 0$.

For a more general version of this theorem, see for example [7]. We give an elementary

Proof. Plugging in the definition of the f-entropy, we get that for all probability distributions $P = (p_i)$, $Q = (q_j)$,

$$\sum_{i,j} f(p_i q_j) = \alpha \sum_i f(p_i) + \beta \sum_j f(q_j) + \gamma \sum_{i,j} f(p_i)f(q_j). \tag{6}$$

Letting Q be deterministic yields $\alpha = 1$ and P deterministic yields $\beta = 1$. For uniform distributions $P = U_m$ and $Q = U_n$, we then find from (6) that

$$f(\frac{1}{mn}) = \frac{1}{m}f(\frac{1}{n}) + \frac{1}{n}f(\frac{1}{m}) + \gamma f(\frac{1}{m})f(\frac{1}{n}), \qquad (7)$$

that is, f satisfies (5) for all $x = \frac{1}{m}$, $y = \frac{1}{n}$. We seek to extend this to all rational numbers in $[0, 1]$, which then extends to all $x, y \in [0, \infty)$ by the analyticity of f.

Let $q \in [0, 1] \cap \mathbb{Q}$ and choose $a, m \in \mathbb{N}$ such that $q + a\frac{1}{m} = 1$. Applying (6) to $P = (q, \frac{1}{m}, \ldots, \frac{1}{m}, 0, \ldots)$ and $Q = U_n$, we get

$$nf(\frac{q}{n}) + anf(\frac{1}{mn}) = f(q) + af(\frac{1}{m}) + nf(\frac{1}{n}) + \gamma\left(f(q) + af(\frac{1}{m})\right)nf(\frac{1}{n}),$$

and so using the above result on the term $anf(\frac{1}{mn})$ and remembering that $\frac{a}{m} = 1 - q$, we get

$$f(\frac{q}{n}) = qf(\frac{1}{n}) + \frac{1}{n}f(q) + \gamma f(q)f(\frac{1}{n}). \qquad (8)$$

Finally, letting also $r \in [0, 1] \cap \mathbb{Q}$ and choosing $b, n \in \mathbb{N}$ such that $r + b\frac{1}{n}$, we apply (6) to $Q = (r, \frac{1}{n}, \ldots, \frac{1}{n}, 0, \ldots)$ and P as above, getting

$$f(qr) + bf(\frac{q}{n}) + af(\frac{r}{m}) + abf(\frac{1}{mn}) = f(q) + af(\frac{1}{m}) + f(r) + bf(\frac{1}{n}) +$$
$$+\gamma\left(f(q) + af(\frac{1}{m})\right)\left(f(r) + bf(\frac{1}{n})\right).$$

Using (8) and (7) on the left-hand terms we get that

$$f(qr) = qf(r) + rf(q) + \gamma f(q)f(r). \qquad (9)$$

Since this holds for all $q, r \in [0, 1] \cap \mathbb{Q}$, the theorem follows by analyticity of f. ////

That this does indeed provide a converse to (3), is shown by the following

Corollary. *If the f-entropy H_f satisfies the conditions of the theorem, then H_f is a Tsallis entropy: $H_f = H_q$ for some $q > 0$.*

Proof. In [1] it is shown that if f satisfies the factorization property

$$\sum_i\left(q_i f(\frac{p_i}{q_i}) - f(p_i)\right) = \sum_i \xi(p_i)\zeta(q_i),$$

where ξ, ζ are analytic functions on $[0, \infty[$ satisfying $\xi(1) = 1$ and $\zeta(1) = 0$, then the f-entropy is a Tsallis entropy.

Well, we have by the theorem that

$$q_i f(\frac{p_i}{q_i}) - f(p_i) = (p_i + \gamma f(p_i))q_i f(\frac{1}{q_i}), \qquad (10)$$

hence the result follows. ////

Before we can tackle the general problem, we need to note an important property of f-entropy. Namely, let $Q = (q_i)_{i \leq N} \in M_+^1(N)^*$ with say, $0 < q_1 < q_2 < 1$, and consider also the distribution $Q_\varepsilon = (q_1 - \varepsilon, q_2 + \varepsilon, q_3, \ldots, q_N)$. If ε is sufficiently small and negative - more precisely, if $\frac{1}{2}(q_1 - q_2) \leq \varepsilon < 0$, we have that the entropy of Q_ε is strictly greater than the entropy of Q. Similarly, if ε is sufficiently small and positive - more precisely, if $0 < \varepsilon \leq q_1$, the entropy of Q_ε will be strictly smaller than that of Q. This follows by the strict convexity of f.

Together with the continuity of H_f on $M_+^1(N)$, this implies that the mapping $\varepsilon \curvearrowright H_f(Q_\varepsilon)$ is a homeomorphism of the interval $[\frac{1}{2}(q_1 - q_2), q_1]$ onto the interval $[H_f(0, q_2 + q_1, q_3, \ldots, q_N), H_f(\frac{1}{2}(q_1 + q_2), \frac{1}{2}(q_1 + q_2), q_3, \ldots, q_N)]$.

Lemma 2. *Let* $0 < a < b < -f'(0)$. *Then there exists* $N \in \mathbb{N}$, *and* $\tilde{q}_1, \tilde{q}_2$ *with* $0 < \tilde{q}_1 < \tilde{q}_2 < 1$ *such that*

$$[a, b] \subseteq \{H_f(Q) | Q = (\tilde{q}_1, \tilde{q}_2, q_3, \ldots, q_N) \in M(\{1, \ldots, N\})\}. \tag{11}$$

The point of this is, that we can fix $N, \tilde{q}_1, \tilde{q}_2$ once and for all, and still have enough probability distributions to "calculate with $g(x, y)$" with $x, y \in]a, b[$.

Proof. Since the sequence $(-nf(\frac{1}{n}))$ is increasing and converges to $-f'(0)$ (indeed, by definition $f'(0) = \lim_{n \to \infty} nf(\frac{1}{n})$), all N sufficiently large will satisfy

$$b < -(N - 1)f(\frac{1}{N - 1}) = H_f(U_{N-1}). \tag{12}$$

So, from the remarks above, it is clear that by choosing N sufficiently large and $\tilde{q}_1 + \tilde{q}_2 = \frac{1}{N-1}$, with $0 < \tilde{q}_1 < \tilde{q}_2$, we get

$$H_f(\tilde{q}_1, \tilde{q}_2, 1 - \tilde{q}_1 - \tilde{q}_2) < a < b < H_f(U_{N-1}) < H_f\left(\tilde{q}_1, \tilde{q}_2, \frac{1}{N - 1}, \ldots, \frac{1}{N - 1}\right). \tag{13}$$

The lemma then follows since we can get the right-hand entropy from the left-hand one by continuously pulling probabilities together, as in the remarks above. ////

Theorem 3. *Assume that the f-entropy H_f satisfies the condition of composability as in eq. (2) for some function g defined on $[0, -f'(0)[\times [0, -f'(0)[$. Then g has partial derivatives on $]0, -f'(0)[\times]0, -f'(0)[$, and indeed*

$$g(x, y) = x + y + \gamma xy, \quad x, y \in [0, -f'(0)[, \tag{14}$$

for some constant γ.

Proof. We want to show that on all sets $]a, b[\times]a, b[$ with $0 < a < b < -f'(0)$, the partial derivatives exist and satisfy

$$\frac{\partial}{\partial x} g(x, y) = A(y), \quad \frac{\partial}{\partial y} g(x, y) = B(x).$$

---

*For $N \in \mathbb{N}$ we write $M_+^1(N)$ for the set of probability distributions on $\{1, \ldots, N\}$.

The theorem then easily follows.

So, for given a, b choose $N, \tilde{q}_1, \tilde{q}_2$ as in the lemma. Let $x, y \in]a, b[$ and choose $P, Q = (\tilde{q}_1, \tilde{q}_2, q_3, \ldots, q_N) \in M_+^1(N)$ such that

$$H_f(P) = x, \quad H_f(Q) = y.$$

If $y_n \to y$ for some sequence (y_n), we can assume[†] that there exists a sequence (ε_n) such that for all n

$$y_n = H_f(Q_{\varepsilon_n}) = (\tilde{q}_1 - \varepsilon_n, \tilde{q}_2 + \varepsilon_n, q_3, \ldots, q_N), \quad \text{and} \quad \lim_{n \to \infty} \varepsilon_n = 0.$$

As

$$
\begin{aligned}
\frac{g(x, y_n) - g(x, y)}{y_n - y} &= \frac{g(H_f(P), H_f(Q_{\varepsilon_n})) - g(H_f(P), H_f(Q))}{H_f(Q_{\varepsilon_n}) - H_f(Q)} \\
&= \frac{H_f(P \otimes Q_{\varepsilon_n}) - H_f(P \otimes Q)}{H_f(Q_{\varepsilon_n}) - H_f(Q)}
\end{aligned}
$$

and as

$$
\begin{aligned}
&\lim_{n \to \infty} \frac{H_f(P \otimes Q_{\varepsilon_n}) - H_f(P \otimes Q)}{H_f(Q_{\varepsilon_n}) - H_f(Q)} \\
&= \lim_{n \to \infty} \frac{\sum_{i=1}^N [f(p_i \tilde{q}_1) - f(p_i(\tilde{q}_1 - \varepsilon_n)) + f(p_i \tilde{q}_2) - f(p_i(\tilde{q}_2 + \varepsilon_n))]}{f(\tilde{q}_1) - f(\tilde{q}_1 - \varepsilon_n) + f(\tilde{q}_2) - f(\tilde{q}_2 + \varepsilon_n)} \\
&= \lim_{n \to \infty} \sum_{i=1}^N p_i \left[\frac{f(p_i \tilde{q}_1 - \varepsilon_n p_i) - f(p_i \tilde{q}_1)}{-\varepsilon_n p_i} - \frac{f(p_i \tilde{q}_2 + \varepsilon_n p_i) - f(p_i \tilde{q}_2)}{\varepsilon_n p_i} \right] \times \\
&\quad \times \left(\frac{f(\tilde{q}_1 - \varepsilon_n) - f(\tilde{q}_1)}{-\varepsilon_n} - \frac{f(\tilde{q}_2 + \varepsilon_n) - f(\tilde{q}_2)}{\varepsilon_n} \right)^{-1} \\
&= (f'(\tilde{q}_1) - f'(\tilde{q}_2))^{-1} \sum_{i=1}^N p_i(f'(p_i \tilde{q}_1) - f'(p_i \tilde{q}_2)),
\end{aligned}
$$

we find that $\frac{\partial}{\partial y} g(x, y)$ exists and that

$$\frac{\partial}{\partial y} g(x, y) = (f'(\tilde{q}_1) - f'(\tilde{q}_2))^{-1} \sum_{i=1}^N p_i(f'(p_i \tilde{q}_1) - f'(p_i \tilde{q}_2)). \tag{15}$$

But since $\tilde{q}_1$ and $\tilde{q}_2$ are the same for any choice of y, it is independent of this choice - i.e.

$$\frac{\partial}{\partial y} g(x, y) = B(x), \quad \text{for all } x, y \in]a, b[. \tag{16}$$

Completely analogously it is shown that the partial derivative with respect to x exists and that

$$\frac{\partial}{\partial x} g(x, y) = A(y), \quad \text{for all } x, y \in]a, b[. \tag{17}$$

[†]By leaving out finitely many of the y_n's if we have to, and using that the mapping $\varepsilon \curvearrowright H_f(Q_\varepsilon)$ is a homeomorphism of intervals as remarked above.

Since a and b were arbitrary, it follows that

$$g(x, y) = C + \alpha x + \beta y + \gamma xy, \quad x, y \in]0, -f'(0)[, \tag{18}$$

and the theorem follows by the continuity of H_f on finite distributions - in particular we have for $P = (p, 1-p)$ and $Q = (q, 1-q)$, $0 < p, q < 1$,

$$
\begin{aligned}
H_f(P) &= \lim_{q \to 1} H_f(P \otimes Q) \\
&= \lim_{q \to 1} [C + \alpha H_f(P) + \beta H_f(Q) + \gamma H_f(P) H_f(Q)] \\
&= C + \alpha H_f(P),
\end{aligned}
$$

so $C = 0$ and (after analogous argument for β) $\alpha = \beta = 1$, which shows the validity of the theorem for all $x, y \in [0, -f'(0)[$. ////

Comparing theorem 3 with the corollary of theorem 1 we see that any f-entropy which is composable is indeed a Tsallis entropy.

3   Discussion

The condition of composability is important in thermodynamical applications of entropy, and we have shown here that any f-entropy which is composable must be a Tsallis entropy. Thus, for example, we see that attempts of 2-parameter generalizations of Tsallis entropy as in [4, 2] as well as many suggestions of alternate entropy functions such as the well known Kaniadikis[‡] entropy will not be able to satisfy the condition of composability.

More generally, entropy is often defined as a function $H(P) = \varphi\left(-\sum_i f(p_i)\right)$, i.e. as a function of f-entropy, where φ is positive and strictly increasing. In particular, φ is bijective, and our theorem shows that if H is composable, then it is in fact a function of some Tsallis entropy. A well known example is the Rényi entropy defined for $\alpha > 0$ by

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left(\sum_i p_i^\alpha \right).$$

It is an interesting question whether there exists one choice of transformation φ which in some sense is better than any other choice. If so, one would expect that either Tsallis or Rényi entropy pops up. For this question, F. Topsøe suggested that his notion of game theoretical equilibrium might be relevant (see [2] for details). In [9] it is shown that the Rényi entropy lacks certain physically important properties that the Tsallis entropy posses.

[‡]more generally, any linear kombination of Tsallis entropies that does not give a new Tsallis entropy.

Acknowledgments

References

[1] F. Topsøe. Factorization and escorting in the game-theoretical approach to non-extensive entropy measures. *Physica A*, 365 (2006), 91-95.

[2] F. Topsøe. Entropy and Equilibrium via Games of Complexity. *Physica A*, 340 (2004), 11-31.

[3] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Physics*, 52 (1988), 479-487, 1988.

   See http://tsallis.cat.cbpf.br/biblio.htm for a complete and updated bibliography.

[4] V. Schwämmle, C. Tsallis. Two-parameter generalization of the logarithm and exponential functions and Boltzmann-Gibbs-Shannon entropy. *arXiv:cond-mat/0703792v1 [cond-mat.stat-mech]*.

[5] J. Havrda, F. Charvát. Quantification method of classification processes. Concept of structural a-entropy. *Kybernetika*, 3 (1967), 30-35.

[6] M. Hotta, I Joichi. Composability and generalized entropy. *Physics Letters A*, 262 (1999), 302-309.

[7] S. Abe. Axioms and uniqueness theorem for Tsallis entropy. *Physics Letters A*, 271 (2000), 74-79.

[8] S. Abe. General Pseudoadditivity of composable entropy prescribed by the existence of equilibrium. *Physical Review E*, 63 (2001), 061105.

[9] S. Abe. Tsallis Entropy: How unique? *arXiv:cond-mat/0305087v1 [cond-mat.stat-mech]*.

[10] R.S. Johal. Additive entropy underlying the general composable entropy prescribed by thermodynamic meta-equilibrium. *Physics Letters A*, 318 (2003), 48-52.

[11] R.S. Johal. Composable entropy and deviation from macroscopic equilibrium. *Physics Letters A*, 332 (2004), 345-349.